HHS Public Access

Author manuscript

Aggress Violent Behav. Author manuscript; available in PMC 2016 January 06.

Published in final edited form as:

Aggress Violent Behav. 2014 ; 19(4): 423–434. doi:10.1016/j.avb.2014.06.008.

A systematic review and content analysis of bullying and cyber-bullying measurement strategies

Alana M. Vivolo-Kantor[a,*], **Brandi N. Martell**[a], **Kristin M. Holland**[a], and **Ruth Westby**[b]

[a]Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, United States

[b]Rollins School of Public Health, Emory University, United States

Abstract

Bullying has emerged as a behavior with deleterious effects on youth; however, prevalence estimates vary based on measurement strategies employed. We conducted a systematic review and content analysis of bullying measurement strategies to gain a better understanding of each strategy including behavioral content. Multiple online databases (i.e., PsychInfo, MedLine, ERIC) were searched to identify measurement strategies published between 1985 and 2012. Included measurement strategies assessed bullying behaviors, were administered to respondents with ages of 12 to 20, were administered in English, and included psychometric data. Each publication was coded independently by two study team members with a pre-set data extraction form, who subsequently met to discuss discrepancies. Forty-one measures were included in the review. A majority used differing terminology; student self-report as primary reporting method; and included verbal forms of bullying in item content. Eleven measures included a definition of bullying, and 13 used the term “bullying” in the measure. Very few definitions or measures captured components of bullying such as repetition, power imbalance, aggression, and intent to harm. Findings demonstrate general inconsistency in measurement strategies on a range of issues, thus, making comparing prevalence rates between measures difficult.

Keywords

Bullying; Scale; Index; Measurement; Instrument

[*]Corresponding author at: Division of Violence Prevention, National Center for Injury Prevention and Control, Centers for Disease Control and Prevention, 4770 Buford Highway NE, MS-F64, Atlanta, GA 30341, United States. Tel.: +1 770 488 1244; fax: +1 770 488 4349. AVivoloKantor@cdc.gov (A.M. Vivolo-Kantor).

Financial disclosure

The authors have no financial relationships relevant to this article to disclose.

Conflict of interest

The authors have no conflicts of interest to disclose.

Disclaimer

All authors certify that the abovementioned manuscript represents valid work and not been submitted for publication elsewhere. The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

Funding source

None.

1. Introduction

Bullying is a form of interpersonal violence that can cause short- and long-term physical, emotional, and social problems among victims, and is, therefore, a serious public health concern (Copeland, Wolke, Angold, & Costello, 2013; Gini & Pozzoli, 2009; Nakamoto & Schwartz, 2009). However, the magnitude of the problem, the prevalence of bullying behavior, the common antecedents of perpetration, and the consequences of bullying are difficult to interpret because the measurement of bullying remains inconsistent among researchers (Atik, 2011; Furlong, Sharkey, Felix, Tanigawa, & Green, 2010). Only recently has the Centers for Disease Control and Prevention (CDC) and the Department of Education (ED) released the first uniform definition of bullying (Gladden, Vivolo-Kantor, Hamburger, & Lumpkin, 2014), however, standardization of bullying measurement[1] is still needed to provide a better understanding of this problem.

Over the past two decades, the concept of bullying has evolved with research. In 1990, Dan Olweus provided a framework for the most widespread contemporary definition of bullying. Specifically, this definition states that bullying includes three key components: intentional aggression, repetition, and a power imbalance (Olweus, 1993). To date, these three components remain a part of the definition of bullying and have been widely used to measure the problem, but they have not been used in a standardized or systematic manner (Grief & Furlong, 2006; Rigby, 2004). There has also been much discussion about additional components that may be required for a behavior to be defined as bullying, including the perpetrator's intent to cause harm and the victim's report of experiencing harm (Greene, 2000; Smith, del barrio, & Tokunaga, 2013; Smith & Thompson, 1991).

CDC and ED's definition of bullying, while similar to previous definitions, introduces some new concepts. Similar to previous definitions, the three main components of unwanted aggressive behavior, observed or perceived power imbalance, and repetition of behaviors are included. However, the definition differs in three ways from other commonly used definitions of bullying. First, the definition requires aggressive behaviors to be *unwanted*. This helps to exclude rough and tumble play among youth. Second, bullying can involve a single act of aggression if it is perceived to have a high likelihood of being repeated (e.g., may involve threats of future aggression). The intent for this inclusion is to encourage timely intervention at the first sign of these behaviors instead of waiting for multiple incidents of aggression to occur. Third, the current definition excludes teen dating and sibling violence. Other bullying definitions do not distinguish teen and sibling violence from peer violence (Gladden et al., 2014).

Researchers have also modified the scope of bullying by incorporating both direct modes of bullying (e.g., fighting) and indirect modes (e.g., rumor spreading), distinguishing between types (e.g., physical, verbal, and relational), and distinguishing similar and sometimes overlapping constructs, such as peer victimization and peer aggression (Farrington, 1993;

[1]The term "measurement strategy" is used in this article to encompass the methods used to assess bullying such as reporter type, time frame, and content of behaviors. When the term "measure" is used, we are specifically describing the items and response options that comprise each scale or index.

Furlong et al., 2010; Olweus, 1993). Most recently, bullying has been adapted to include "cyber-bullying," a form of inter-net and electronic harassment (Tokunaga, 2010).

Although research has provided unique perspectives about bullying and improved understandings of the nature of this form of violence, significant inconsistencies still remain with respect to bullying definitions and measurement strategies currently used in studies. These inconsistencies can provide confiicting prevalence estimates and scientific results. A review and meta-analysis of bullying in school-based studies using varying measurement strategies concluded that 53% of students, on average, reported exposure to bullying (as victims, bullies, or bully/victims). However, prevalence ranges drastically for each category; for bullying perpetration, the range was 5% to 44% (Cook, Williams, Guerra, & Kim, 2010). Inconsistent measurement strategies can also increase the difficulty in monitoring the problem through public health surveillance initiatives and evaluating the impact and progress of public health bullying prevention interventions. For example, the Youth Risk Behavior Survey (YRBS) and the School Crime Supplement to National Crime Victimization Survey (NCVS) measure bullying and cyberbullying in drastically different ways. The 2013 YRBS bullying questions starts with "*Bullying is when 1 or more students tease, threaten, spread rumors about, hit, shove, or hurt another student over and over again. It is not bullying when 2 students of about the same strength or power argue or fight or tease each other in a friendly way. During the past 12 months, have you ever been bullied on school property*?" and "*During the past 12 months, have you ever been electronically bullied*? (*Include being bullied through e-mail, chat rooms, instant messaging, Web sites, or texting*.)" Using response options of "yes" or "no", CDC found that 19.6% reported being bullied and 14.8% reported being electronically bullied (Kann et al., 2014). Yet, using the NCVS with a similar age group and time frame, the 2009 results demonstrated larger prevalence rates for school bullying victimization (32%) and smaller prevalence for electronic bullying (4%) (DeVoe & Murphy, 2011). To measure bullying, the NCVS asks youth, "*Now I have some questions about what students do at school that make you feel bad or are hurtful to you. We often refer to this as being bullied. You may include events you told me about already. During this school year, has any student bullied you*? *That is, has another student*…" followed by a set of seven behavioral questions (i.e., made fun of you, called you names, or insulted you; spread rumors about you; threatened you with harm).

1.1. Goal of the current study

This review identifies various measurement strategies of bullying behaviors among youth and provides suggestions for standardizing measurement for research and surveillance purposes. Specifically, we examined youth bullying measurement strategies to identify variations in data collection methods, terminology used, and definitional components. We provide an overview of reliable measures that identify specific components of bullying (e.g., power imbalance, intention to harm, and repetition) and the content of bullying behaviors (e.g., hitting, kicking, rumor spreading). We also provide an overall assessment of bullying measurement strategies, indicate the usefulness of acknowledged measures, and recognize the next steps to establish a well-constructed measure of bullying behaviors.

2. Method

2.1. Literature search

A systematic search was conducted for all bullying and cyber-bullying measurement strategies published between 1985 and 2012. First, key search terms were drawn from a review of the literature and included such terms as bully*, violen*, aggress*, victim*, harass*, exclude*, bystand*, measure*, tool*, and survey*. The search terms were used in combination with each other to narrow the search results. For example, the terms “bullying”, “victimization” and “survey” were entered simultaneously to retrieve relevant publications. For a full list of included search terms, please contact the corresponding author of this article.

Using these terms, a search was performed of the following electronic databases: PsychInfo, PsychArticles, MedLine, ERIC, the Psychology and Behavioral Sciences Collection, the Professional Development Collection, SocIndex with Full Text, Expanded Academic Index ASAP, and Science Direct. The abstracts of all relevant articles were screened for inclusion eligibility. When there was sufficient indication that a publication abstract was appropriate for consideration, the publication was retrieved, the full publication was reviewed, and the measure was obtained by contacting authors or copyright holders if it was not available within the publication.

In addition to the search described above, we also included measurement strategies published in the CDC document, “*Measuring Bullying Victimization, Perpetration, and Bystander Experiences: A Compendium of Assessment Tools*” (also known as the *Bullying Compendium*) (Hamburger, Basile, & Vivolo, 2011). The *Bullying Compendium* represents a comprehensive list of bullying measurement strategies used by researchers in the field; however, the Compendium does not provide an in-depth overview of the measures. The current paper, on the other hand, provides a detailed review of constructs measured and definitions used for each bullying measurement strategy, and also identifies advantages and drawbacks of each in an attempt to more consistently guide research efforts.

2.2. Inclusion and exclusion criteria

Measurement strategies included in this study were ones that: (a) assessed *bullying behaviors*, including physical/psychological bullying and victimization, cyber-bullying/electronic aggression, relational aggression, sexualized and homophobic bullying, and bullying bystander behaviors; (b) was administered to respondents between, but not limited to, age 12 and 20, or administered to parents, teachers, peers, or other individuals who could report on the behaviors of youth within this age range; (c) was developed or revised and examined between 1985 and 2012; and (d) was administered in English. We also reviewed measurement strategies described in published peer-reviewed and non-peer reviewed journals, book chapters, and online sources. Regardless of how the measurement strategy was published, the final inclusion criterion was that (e) the publication included psychometric data. Studies were excluded if they did not meet the abovementioned criteria or the measurement strategy: (a) only assessed beliefs, attitudes, or perceptions; (b) was not a minor adaptation of included measures; (c) did not include psychometric data; (d) was not

a scale or index; or (e) explicitly explored workplace bullying. We also excluded measures if we (f) could not retrieve the publication for review after contacting the developer(s).

Minor adaptations were defined as any adaptation that kept intact the response options, time frame, and content of the items. In most cases, minor adaptations included decreasing the number of items. In several situations, measures were shortened to improve reliability and validity. In these situations, both measures were included and reviewed. Also, this review focuses on measures that can be defined as scales or indices. Scales are defined as multiple items that measure only one concept using the same or similar response options, while indices include multiple questions with different response options that may not be correlated with other index items (Bollen & Lennox, 1991; DiIorio, 2006; Salazar, DiClemente, & Crosby, 2011). We have excluded one-item scales because these typically lack precision, may change over time, and are narrowly defined (Spector, 1992). One-item measures of bullying victimization and perpetration typically involve assessing whether an individual has ever experienced a behavior. While this may be relevant information for assessing bullying prevalence, these measures do not provide a thorough understanding of bullying behaviors.

2.3. Search results

The search began in October, 2012 and concluded in April, 2013. Over 1000 publications were screened, and a total of 164 bullying measures were deemed relevant for abstract screening. Following abstract screening, the publications for 69 measurement strategies were retrieved and reviewed for eligibility. Twenty-eight were subsequently excluded because they failed to meet our inclusion criteria. Forty-one publications met the inclusion criteria and were included in this review. Fig. 1 provides detailed information regarding reasons for publication exclusion.

2.4. Data extraction and coding process

2.4.1. Data extraction form—A data extraction form was developed to capture all information required to complete this review. The standardized form included 32 questions covering a range of information. The form included two items on publication information: publication type (e.g., journal versus report) and publication year. We collected data on sample size characteristics, including the total number of participants in the sample, the target age range, and the target grade level. The form was also used to collect information about the measurement strategy implementation setting (e.g., school-based survey, clinic-based survey) and location (e.g., domestic vs. international).

The field currently uses an array of terminology and definitions to describe bullying (Swearer, Siebecker, Johnsen-Frerichs, & Wang, 2010). The extraction form captured information about how the authors described bullying in each publication. For example, some publications used the term “peer victimization” or “peer aggression” when describing bullying. Our form also captured information on whether participants were provided a specific definition of bullying and if so, what information the definition encompassed. Because little consensus exists among researchers regarding which components should be included in a bullying definition (Gladden et al., 2014), we noted which of the five most

commonly used components, if any, were included (i.e., power imbalance, repetition, intent to harm, victim experiences harm, and intentional aggressive behaviors).

The reliability and validity of each measure were recorded. In situations where the primary publication used for coding referenced a different publication when discussing psychometric data, the study team retrieved and coded that publication only for psychometric data.

We captured data specific to each measurement strategy such as the author-reported type(s) of bullying included in the measure (i.e., relational, verbal, physical, indirect, and/or direct). We also noted the reporter for each strategy (e.g., youth self-report, peer nomination), the number of items in the measure, the response options (e.g., yes/no, Likert scale), the time frame assessed (e.g., past week), whether the measurement strategy specifically included the term “bullying”, how the measure was scored, and if the measure captured victimization, perpetration, or bystander experiences.

Because of potential differences between the manner in which the author defined bullying and the measures administered, we captured data on the components that were measured. For example, if a measure assessed repetition, a description of how repetition was measured (e.g., one incident occurring over a span of time, more than one incident involving the same perpetrator or group of perpetrators) was subsequently recorded. Similarly, we specified how power imbalance was measured (e.g., perpetrator(s) has more physical strength, multiple perpetrators, perpetrator(s) is more popular). In addition, we captured data on the content of behaviors assessed by each measure such as hitting, kicking, throwing objects, destructing property, spreading rumors, name calling, homophobic teasing, and threatening.

2.4.2. Coding process—Prior to coding all publications, all four members of the study team independently coded the same publication. After confirming high levels of agreement across coders and coding form questions using Cohen’s kappa (range per question $\kappa = 0.82–1.00$) (Landis & Koch, 1977), each publication was coded independently by two study team members. The two coders assigned to each publication subsequently met to discuss any discrepancies. When two coders did not agree on a code, the other team members were consulted and provided recommendations, so that a consensus was reached among all team members.

2.5. Data analyses

IBM SPSS Statistics 21 was used to conduct analyses focused on characterizing the features of included measurement strategies with frequencies and other descriptive statistics. In several situations, specifically those requiring analysis of responses to open-ended questions from the data extraction form, new variables were created to re-code the data extracted from the measurement strategies.

To best capture pertinent information about the content of bullying behaviors assessed by each measure, a summative, or manifest, content analysis (Potter & Levine-Donnerstein, 1999) was completed, where the appearance of particular content within the items (e.g., hit, kick, or push) was coded and grouped to include all similar contents in each measure. Using the example provided above, all items that included hitting, kicking, or pushing of another

youth were combined to form a content area of “physical bullying”. This process was used to determine all relevant bullying behavior contents. A total of 17 behavior content categories were identiied.

3. Results

3.1. Measurement strategy characteristics

Of the 41 included measurement strategies, most were published in peer-reviewed journal articles (n = 39, 95.1%) between 1988 and 2012, with the majority published after 2003 (n = 27; 65.8%). Measures were administered among samples within the United States (n = 19, 46.3%) and internationally (n = 15, 36.6%), as well as in multiple countries simultaneously (n = 3, 7.3%). The mean sample size was 1089 (SD = 1638; range, 47–8693). Almost all measurement strategies were implemented in schools (n = 38, 92.8%), but several were implemented in other settings such as prisons (n = 1, 2.4%) and by mail (n = 1, 2.4%). One setting (2.4%) was unknown. The included measures were administered among youth between the ages of 3 and 25; the average age range was 10.59–15.73 years. Table 1 provides more information on measure characteristics.

3.1.1. Terminology and types of bullying—The use of varying terminology by authors was captured in this review. The majority described bullying as the behavior(s) assessed by the measure (n = 29, 70.7%); however, a notable proportion used the terms “peer victimization” (n = 14, 34.1%) and “peer aggression” (n = 12, 29.3%) to describe the behaviors they measured. About a third of the strategies used the term bullying either as a precursor to the measure (e.g., “How did you get bullied?” followed by several behavioral items) (Swearer & Cary, 2003) or as an item within the measure (e.g., “I was bullied at school,” [(Hinduja & Patchin, 2010)]) (n = 13, 31.7%). In most cases, youth self-report of bullying behavior was the primary assessment method (n = 35, 85.4%), followed by peer nomination (n = 9, 22%). Occasionally (n = 4, 9.7%), self-report and peer nomination were used together. In two measures, self-report, peer nomination, and teacher-report were used in conjunction.

The author-reported type(s) of bullying were captured for 38 of the 41 measures (7.3% were unknown). Most often, authors described the inclusion of items assessing verbal forms of bullying (n = 34, 82.9%), followed by direct bullying (n = 30, 73.2), physical bullying (n = 29, 70.7%), relational bullying (n = 22, 53.7%), indirect bullying (n = 17, 41.5%), and cyber-bullying (n = 7, 17.1%). For 29 measures (70.7%), the authors stated that physical and verbal bullying were both included, while no authors said physical, verbal, and relational bullying were all included in the measurement strategy. However, 13 (31.7%) authors stated that physical, verbal, and indirect bullying were measured.

3.1.2. Assessment of behavioral content—The content of the items included in the measures also varied drastically across publications. While we captured the author-stated type of bullying being measured as mentioned earlier (i.e., physical, verbal, relational, cyber, indirect, direct, other), a more in-depth content analysis of the measures’ items was conducted to accurately categorize the behavioral content. Table 2 lists the 17 behavioral content categories developed by analyzing the items in all measures and reports the

measures that assessed each construct. Many measures included items on making fun/teasing/embarrassing others (n = 33, 80.5%), physical bullying (n = 32, 78%), specific name calling (n = 30, 73.2%), threatening others (n = 25, 61%), socially excluding others from groups or activities (n = 23, 56.1%), and making rude comments and gestures (n = 22, 53.7%).

3.1.3. Use of definitions and measuring definitional components—Eleven measurement strategies (26.8%) included a definition of bullying. Among those with definitions, less than half (n = 4, 36.4%) captured all five components (i.e., power imbalance, intention to harm, victim experiences harm, repetition, and aggressive behaviors) recommended by the field (Gladden et al., 2014; Olweus, 1993) for inclusion in a bullying definition. Three (27.2%) captured four components, and four (36.4%) captured three or less. The components most often included in the definition were power imbalance (n = 9, 81.8%), intention to cause harm (n = 9, 81.8%), aggressive behaviors (n = 8, 72.7%), repetition (n = 7, 63.6%), and victim experiences harm (n = 6, 54.5%).

Because most measures did not include an explicit definition of bullying, we identified which components of the recommended bullying definition were assessed by the measure (e.g., inclusion of items or response options that allowed for a power imbalance to be measured). Regardless of whether a definition was provided, items measuring aggressive behaviors were included most often (n = 39, 95.1%), followed by repetition (n = 32, 78%), intention to harm (n = 26, 63.4%), victim experiences harm (n = 16, 39%), and power imbalance (n = 9, 22%). When segregating the data by those measures with and without a definition, Fisher's exact test yielded no significant differences at the $p = 0.01$-level for the definitional components actually measured in the scale or index.

For the 32 measures where repetition was denoted, the study team determined how repetition was measured. In about half of these measures (n = 17, 53.1%), authors used broad frequency response options for each behavioral item such as "How often have you taken things from other students?" with response choices "never", "sometimes", and "often" (Raine et al., 2006). In the rest of the measures (n = 15, 46.9%), repetition was assessed using the actual number of times the incident occurred (e.g., "Some kids call each other names such as gay, lesbo, fag, etc. How many times in the last week did you say these things to a friend?" Response options included never, 1 or 2 times, 3 or 4 times, 5 or 6 times, and 7 or more times) (Poteat & Espelage, 2005).

When a power imbalance was denoted in the measurement strategy (n = 9), the same process as was used for repetition was implemented. Most often (n = 5, 55.6%) items included mention of the perpetrators' physical strength (e.g., "Please think about the main person or leader who did these things to you in the past month. How physically strong is this student?" Response options included "less than me", "same as me", and "more than me") (Felix, Sharkey, Green, Furlong, & Tanigawa, 2011), followed by items describing multiple perpetrators (n = 3, 33.3%) (e.g., "In the past month a group of kids tried to beat me up." Response options included never, once or twice, three or four times, and five or more times) (Peters & Bain, 2011), perpetrators who were older/in a higher grade (n = 3, 33.3%), perpetrators who were more popular (n = 1, 11.1%), perpetrators who were adults (n = 1,

11.1%), and perpetrators who were smarter (n = 1, 11.1%). Table 3 provides additional information about the measures that included definitions, the components included in the definition, and a detailed breakdown of these components.

3.2. Scoring strategies

Scoring strategies for each measure varied by publication. However, in over half of the measures (n = 21, 51.2%), responses were summed to yield a total score for the overall scale/index or subscale. This summed score was then used as a continuous outcome variable where higher scores were predictive of higher levels of perpetration, victimization, or bystander experiences. Eleven measures (26.8%) classified bullying into binary categories by either summing across responses and dichotomizing based on “never” versus “ever” or creating binary categories by using a cut-off score. For example, the Traditional Bullying and Cyber-bullying Scale (Hinduja & Patchin, 2010) creates a summed score for each subscale (e.g., bullying victimization, bullying perpetration, cyber-bullying victimization, and cyber-bullying perpetration) and then dichotomizes each subscale into “never/once or twice” to denote no or low frequency of bullying versus “three or more times” to denote higher frequency of bullying. In another example, the California Bully Victimization Scale (Felix et al., 2011) categorized participants into “bullied victims” by using a cut-off score where youths were classified as bullied victims if they reported victimization of one type of bullying (i.e., teasing) at least 2–3 times a month or more and endorsed at least one type of power imbalance. Scoring for measures that included peer nomination (n = 5, 12.2%) was mostly determined by calculating an individual score for each youth nomination and summing across all categories (i.e., victimization or perpetration). The Participant Role Questionnaire (Salmivalli & Voeten, 2004) computes youths’ peer-evaluated sum scores on each of the five subscales and divides by the number of peer evaluators, which produces a continuous score from 0 (never) to 2 (a lot) for each student on each subscale. Scores range from 0 to 24 for victimization and 0 to 20 for perpetration, with higher scores indicating more experiences as a victim or bully.

3.3. Validity and reliability

All included measurement strategies reported validity and/or reliability statistics. However, not all strategies reported the same types of statistics. Specifically, 13 (31.7%) reported several types of validity tests such as face; construct (e.g., convergent and discriminant); and criterion validity (e.g., concurrent and predictive). For construct validity, scales were compared to other bullying scales such as the Olweus Bully/Victim Questionnaire (Solberg & Olweus, 2003) or The Swearer Bully Survey (Swearer & Cary, 2003); teacher assessments and observations; and other student self-report measures such as prosocial behaviors or other behavioral or attitudinal predictor variables. Other scales, for example the Multidimensional Peer-Victimization Scale (Mynard & Joseph, 2000), compared several bullying-related questions to ensure convergent validity. Youth were first asked to self-report victimization by answering yes or no to the question, “Have you ever been bullied?” and were grouped into “victims” and “non-victims.” Then youths responded to 16 behaviorally-based items that began with, “How often during the last school year has another pupil done these things to you?” Specific items included called me names, punched me, and made other people not talk to me. Comparisons found convergent validity with

significant mean differences on self-reports of being bullied between victims and non-victims among all four main factors—physical victimization, verbal victimization, social manipulation, and attacks on property.

Of the 41 measurement strategies, 37 (90.2%) reported Cronbach's alpha for internal consistency, 11 (26.8%) reported test–retest reliability statistics, and one (2.4%) reported split-half reliability. See Table 3 for details on the reliability of each measure.

3.3.1. Internal consistency—Internal consistency ranged from $\alpha = 0.25–0.96$ (mean = 0.82) for overall bullying perpetration and $\alpha = 0.68–0.97$ (mean = 0.84) for overall bullying victimization.

3.3.2. Test–retest reliability—Only one measure reported an overall test–retest (i.e., perpetration and victimization combined) correlation; $r = 0.79$ when youth ages 10–13 were surveyed two weeks apart. Test–retest correlations for victimization ranged from $r = 0.61–0.94$ (mean = 0.82) and perpetration ranged from $r = 0.76–0.90$ (mean = 0.83).

3.3.3. Split-half reliability—One measure reported split-half reliability, where the measure was split into two sections and scores for each section were compared to determine consistency in measurement. These correlations ranged from $r = 0.55$ to $r = 0.82$.

4. Discussion

The aim of the current study was to conduct a systematic review and content analysis of bullying measures administered to youth, teachers, and parents in an effort to gain a better understanding of the strategies employed and the specific components of bullying being measured. Findings suggest that there are important discrepancies between bullying measurement strategies, such as the time frame used to assess when bullying occurred, the components included in bullying definitions, and the behavioral content of measures provided to participants. Of the 41 measures included in this review, most were implemented in school settings, and very few measured bullying occurring outside of schools or in homes. Cyber-bullying, which has traditionally been viewed as an issue not addressed by schools, was not assessed by most of the measures included in this study.

The most predominant method used to assess bullying was youth self-report. While self-report has been the most widely used method, many have suggested that challenges exist in using this method as the sole strategy to collect information on an individual's behavior (Furlong, Sharkey, Bates, & Smith, 2004; Leff, Power, & Goldstein, 2004). Because it is important to achieve the most accurate assessment of the frequency and magnitude of these behaviors, multiple methods should be considered. For instance, prevalence estimates may increase as the awareness of what constitutes bullying increases, thus self-report alone may not be sensitive enough to detect real changes in the rate of bullying. In addition, the field knows very little about the accuracy of self-report bullying measurement. Only a handful of studies have introduced peer nomination, school records, or parent report to supplement information gleaned from youth self-report (Cornell & Brockenbrough, 2004; Leff et al., 2011), and only four measures in this review used multiple reporters to assess bullying. In

fact, research by Cornell and Brockenbrough (2004) found very low agreement among student self-report, peer nomination, and teacher nomination of students as victims of bullying in a rural sample of middle school youth. Interestingly, they found better agreement between peer nomination and teacher nomination at identifying both bullying perpetrators and victims. Because this research raises concern about the sole use of student self-report measurement methods, future research should aim to implement multiple-source reporting to assess bullying behaviors with a national sample of youth.

Regardless of reporting method, almost all of the measures in this review captured both victimization and perpetration of bullying. With increasing evidence that youth are often both victims and perpetrators, it is important to continue to capture both behaviors in measurement. These individuals, also called “bully/victims,” report negative outcomes as much as, if not more than, individuals who are only victims or only perpetrators (Haynie et al., 2001; Nansel, Craig, Overpeck, Saluja, & Ruan, 2004; Veenstra et al., 2005). In this review, few measures included items to better understand bystanding or witnessing behaviors even though there is mounting evidence that bystanders or witnesses also experience similar deleterious effects of bullying (Nishina & Juvonen, 2005; Rivers, Poteat, Noret, & Ashurst, 2009). Without capturing victimization, perpetration, and bystanding behavior measurement, it could be more difficult to target interventions for those most at risk.

The field of bullying is in desperate need of uniform terminology and definitions to describe these behaviors (Swearer et al., 2010; Vivolo, Holt, & Massetti, 2011). In this review, authors used several terms to discuss bullying behaviors, including peer victimization and peer aggression. The use of inconsistent terminology is problematic for several reasons. First, specific to peer victimization and peer aggression, the term “peer” denotes someone of equal status, age, or grade. However, one of the key constructs in bullying definitions, as mentioned earlier, is the presence of a power differential or imbalance in the relationship between the victim and the perpetrator. Thus, this terminology is in direct conflict with the construct of bullying.

Second, there is a growing literature that distinguishes aggression, fighting, and even legally-defined harassment, from bullying (Crick & Dodge, 1996). Many researchers have developed subscales that measure fighting separate from bullying. For example, analyses of the Illinois Bully Scale (Espelage & Holt, 2001; Espelage, Low, Rao, Hong, & Little, 2013) have demonstrated using confirmatory factor analysis that the items capturing physical fighting result in a subscale different from bullying, which included behaviors such as teasing other students, upsetting other students for the fun of it, excluding others from their group of friends, helping to harass other students, and threatening to hit or hurt another student. It is possible that these forms of aggression and violence differ by perceived reasoning and decision processes. Research by Crick and Dodge (1996) interpreted the differences between children who use proactive (i.e., a deliberate behavior to obtain a desired goal) and reactive (i.e., a response driven by anger, frustration, or provocation) aggression. Additionally, there is some evidence that interventions that target physical fighting and other forms of aggression or youth violence are unsuccessful in preventing bullying behaviors (Espelage, Low, Polanin, & Brown, 2013; Taub, 2002; Van Schoiack-

Edstrom, Frey, & Beland, 2002), and some bullying prevention programs are not effective at preventing violence and aggression (Ferguson, San Miguel, Kilburn, & Sanchez, 2007). This illustrates the need to address bullying as a distinct construct that should be examined separately from physical fighting and aggression that is neither repeated, nor involves a power imbalance.

Further, this review uncovered 13 measures that included the term bullying in their strategy and 11 that included a bullying definition. Using the term bullying without providing additional guidance for youth in the form of a definition or list of behaviors may be problematic, as research is mixed on how youth perceive this term. Smith, Cowie, Olafsson, and Liefooghe (2002) found that terms like “bullying” and “picking on” clustered together while terms such as “harassment” and “intimidation” fell into a separate cluster; thus, these terms are not always synonymous with bullying. Using the term bullying in measurement may also impact prevalence. Results from Kert, Codding, Tryon, and Shiyko (2010) show that youth reported significantly less bullying behavior when the word “bully” was provided in a measure than youth not provided a measure with the term included. Another problem with using the term “bullying” in measurement is the fact that recent research suggests that youths may not perceive bullying as researchers do. Land (2003) presented several terms to students (i.e., teasing, bullying, and sexual harassment) and asked them to provide examples of what constituted these terms. The author found that a key component of bullying (e.g., repetition) was not included in students’ examples. Similarly, research from Vaillancourt et al. (2008) revealed discrepancies with respect to youths not including power imbalance and intentionality in a definition of bullying.

Skepticism around using a researcher-developed definition without explicit examples of bullying behaviors has increased over time. Vaillancourt et al. (2008) found that students who were given a definition of bullying in measurement reported less victimization than students who were not provided a definition. This finding has important implications for establishing accurate prevalence rates of bullying. In the current review, the components most often included in the definitions of bullying that prefaced bullying measures were power imbalance, intention to cause harm, and aggressive behaviors, and only four measures included all five components of bullying as recognized by experts in the field. The exclusion of specific components of bullying behaviors from measures of bullying calls into question the validity of the construct being measured.

The vast majority of publications included in this review did not provide a definition in their measurement strategy. Without a presented definition it remains unclear what the author’s a priori definition of bullying included. The current review provides a thorough examination of how measurement strategies with and without explicit definitions integrate the definitional components of bullying and can be used as a guide for bullying researchers in their plans for measuring bullying-related behaviors. Because bullying has been used as a catch-all phrase to encompass a broad category of behaviors (i.e., physical, verbal, relational), Cornell, Sheras, and Cole (2006) have asked “whether all these forms of bullying are psychologically equivalent.” For example, several definitions include multiple behaviors such as hitting, teasing, and spreading rumors; however, incorporating these behaviors in the same definition may fail to capture the nuances associated with each type of behavior.

The time frames used for reporting also vary drastically based on measurement strategy. In fact, most of the measures included in the review did not provide a specified time range, which presents problems in terms of both comparing bullying rates within the sample of interest and between multiple samples. Among those that did provide a time frame, there were variations in the time frames assessed (e.g., nine used the time frame “past 30 days”, four used “past 7 days/week”, two used “current school year”). Although these variations in reporting periods may seem slight, the differences can result in great disparities in prevalence estimates, particularly depending on the time of year when measures are administered. For example, measures administered in September with the time frame “past 30 days” may have students reporting on behaviors that occurred over the summer; a similar measure administered in February would likely yield different results, given that school is in session in the month of January and therefore the opportunity for bullying perpetration/victimization is theoretically higher. Further, four measurement strategies instructed youths to indicate the frequency with which behaviors occurred with respect to other time frames, such as “recently”. Estimating prevalence rates using such broad terms can be problematic since individuals may interpret the meanings of such terms differently. Overall, the differences in reporting time frames make comparing prevalence rates between samples difficult, if not impossible.

Almost all of the included measures provided Likert-type response options, less than half used binary response options (e.g., yes/no, true/false) or open-ended questions, and two used multiple choice responses to assess bullying behaviors. The variation in response options likely impacts not only overall prevalence rates, but also the kind of information being reported. For instance, responses to open-ended questions may garner more or less detail about bullying behaviors, depending on the extent to which the respondent elaborates. Again, based on the various response options used in different measurement strategies, comparing prevalence rates of bullying overall, or even specific components of bullying behavior, becomes nearly impossible, as there is no clear way to draw parallels between behaviors that occur, for instance, “frequently” as judged by a 5-point Likert-type response option to those that have occurred at least once as judged by a “yes” response to a binary item.

Finally, in terms of scoring the measures that assess bullying behaviors, it is difficult to synthesize results across measures. Most often, measures were summed to yield a total score and then characterized as continuous. Eleven measures created binary categories by either summing across responses and dichotomizing based on “never” versus “ever” or creating binary categories by using a cut-off score. Five used peer nomination strategies by which youths identified peers as victims, bullies, and/or bully/victims. The characterization of bullying in a sample greatly depends on how measures are scored and on how bullies and victims are identified. Thus, scoring techniques are critical in establishing accurate estimates of bullying rates.

In addition to the variability in establishing solid prevalence estimates based on the criteria described earlier, most bullying measurement strategies used in the field lack sufficient psychometric properties, including reliability and validity. These characteristics are fundamental to accurately assessing bullying prevalence. In this review, almost all studies

reported moderate to high reliability for their measures, indicating that they consistently yielded similar findings, but most did not assess the validity of the measure, thus leaving the question of whether the surveys accurately measured what they aimed to measure unanswered. Those that did assess validity mostly reported low convergent, discriminant, concurrent, and predictive validity, suggesting that these measures did not clearly assess the intended construct(s). While it is important for measures to be reliable so that researchers can be sure that they are consistently measuring their construct of interest, reliability does not imply validity. It is critical that researchers aiming to assess bullying behaviors are accurately measuring those behaviors, not only in terms readily interpreted by the researchers themselves, but also in terms that recognize the differing perspectives of the youths being surveyed. Thus, in future implementation of measurement strategies, researchers should determine both the reliability and validity of their measures.

4.1. Research limitations

There are several limitations of this review. One particular limitation is that several measures were excluded due to developer non-response to repeated emails for additional information. It is possible that the measures not included based on this factor may be different from the ones included, thus affecting our results. Second, the publications with enough information to describe the measurement strategy tended not to include prevalence data, and therefore we were unable to analyze the data on prevalence by measurement strategy. Next steps should determine how to capture best the relevant prevalence or incidence data to make comparisons across measurement strategies. Lastly, very few measures meeting our inclusion criteria included cyber-bullying items or were dedicated solely to measuring cyber-bullying behavior. A systematic review of cyber-bullying measurement conducted by Berne et al. (2013) included only measures that assessed web-based or electronic bullying behaviors. The authors found that very few cyber-bullying measures stated that their aim was to measure bullying, nor were most measures assessed for reliability and validity. Additional research is needed to better integrate cyber-bullying measures with traditional bullying measurement. Despite these limitations, the results of this study still provide important information about the measures currently being used to assess bullying behaviors, including the measurement strategies employed and the behavioral content assessed by the measures.

4.2. Conclusion

There is much inconsistency in the manner in which bullying is measured by researchers. These inconsistencies range from differences in terminology and temporal referent period to differences in definitional components and actual behaviors measured by the surveys. While these inconsistencies may seem minor, they most likely explain the wide variation in bullying prevalence rates obtained by researchers in the field. Our results further highlight the need for a consistent definition of bullying, which has major implications for the measurement of the construct and the prevention of its occurrence. Future research should focus on integrating a honed definition of bullying into the development of new or improved measurement strategies so that bullying can be more accurately and precisely assessed.

Acknowledgments

The authors would like to thank Dr. Laura Salazar for contributing feedback on drafts of the manuscript.

References

Atik G. Assessment of school bullying in Turkey: A critical review of self-report instruments. Procedia Social and Behavioral Sciences. 2011; 15:3232–3238. http://dx.doi.org/10.1016/j.sbspro.2011.04.277.

Austin S, Joseph S. Assessment of bully/victim problems in 8 to 11 year-olds. British Journal of Educational Psychology. 1996; 66(4):447–456. http://dx.doi.org/10.1111/j.2044-8279.1996.tb01211.x. [PubMed: 9008423]

Berne S, Frisén A, Schultze-Krumbholz A, Scheithauer H, Naruskov K, Luik P, et al. Cyberbullying assessment instruments: A systematic review. Aggression and violent behavior. 2013; 18:320–334. http://dx.doi.org/10.1016/j.avb.2012.11.022.

Bollen K, Lennox R. Conventional wisdom on measurement: A structural equation perspective. Psychological bulletin. 1991; 110(2):305. http://dx.doi.org/10.1037//0033-2909.110.2.305.

Bond L, Wolfe S, Tollit M, Butler H, Patton G. A comparison of the Gatehouse bullying scale and the peer relations questionnaire for students in secondary school. Journal of School Health. 2007; 77(2):75–79. http://dx.doi.org/10.1111/j.1746-1561.2007.00170.x. [PubMed: 17222158]

Bosworth K, Espelage DL, Simon TR. Factors associated with bullying behavior in middle school students. Journal of Early Adolescence. 1999; 19:341–362. http://dx.doi.org/10.1177/0272431699019003003.

Cerezo F, Ato M. Bullying in Spanish and English pupils: A sociometric perspective using the BULL-S questionnaire. Educational Psychology. 2005; 25:353–367. http://dx.doi.org/10.1080/01443410500041458.

Cetin B, Yaman E, Peker A. Cyber victim and bullying scale: A study of validity and reliability. Computers & Education. 2011; 57(4):2261–2271.

Chan JHF, Myron RR, Crawshaw CM. The efficacy of non-anonymous measures of bullying. School Psychology International. 2005; 26:443–458. http://dx.doi.org/10.1177/0143034305059020.

Cho J, Hendrickson J, Mock D. Bullying status and behavior patterns of preadolescents and adolescents with behavioral disorders. Education and Treatment of Children. 2009; 32(4):655–671.

Cook, CR.; Williams, KR.; Guerra, NG.; Kim, TE. Variability in the prevalance of bullying and victimization: A cross-national and methodological analysis. In: Jimerson, SR.; Swearer, SB.; Espelage, DL., editors. Handbook of bullying in schools: An international perspective. New York: NY: Routledge; 2010. p. 347-362.

Copeland, WE.; Wolke, D.; Angold, A.; Costello, EJ. Adult psychiatric outcomes of bullying and being bullied by peers in childhood and adolescence. JAMA Psychiatry. 2013. http://dx.doi.org/10.1001/jamapsychiatry.2013.5041-8

Cornell DG, Brockenbrough K. Identification of bullies and victims: A comparison of methods. Journal of School Violence. 2004; 3(2–3):63–87. http://dx.doi.org/10.1300/J202v03n02_05.

Cornell, DG.; Sheras, PL.; Cole, JCM. Assessment of bullying. In: Jimerson, S.; Furlong, M., editors. Handbook of school violence and school safety: From research to practice. Mahwah, NJ: Erlbaum; 2006. p. 191-210.

Crick NR, Dodge KA. Social information-processing mechanisms in reactive and proactive aggression. Child development. 1996; 67(3):993–1002. http://dx.doi.org/10.2307/1131875. [PubMed: 8706540]

Crick NR, Grotpeter JK. Relational aggression, gender, and social psychological adjustment. Child Development. 1995; 66:710–722. http://dx.doi.org/10.2307/1131945. [PubMed: 7789197]

DeVoe, J.; Murphy, C. Student reports of bullying and cyber-bullying: Results from the 2009 School Crime Supplement to the National Crime Victimization Survey. National Center for Education Statistics; 2011. p. 52

DiIorio, CK. Measurement in health behavior: Methods for research and evaluation. Vol. 1. San Francisco, CA: Wiley; 2006.

Eisenberg ME, Neumark-Sztainer D, Perry CL. Peer harassment, school connectedness, and academic achievement. Journal of School Health. 2003; 73:311–316. http://dx.doi.org/10.1111/j.1746-1561.2003.tb06588.x. [PubMed: 14593947]

Espelage DL, Holt MK. Bullying and victimization during early adolescence. Journal of Emotional Abuse. 2001; 2(2–3):123–142. http://dx.doi.org/10.1300/J135v02n02_08.

Espelage, DL.; Low, S.; Polanin, JR.; Brown, EC. The impact of a middle school program to reduce aggression, victimization, and sexual violence. Journal of Adolescent Health. 2013. http://dx.doi.org/10.1016/j.jadohealth.2013.02.021

Espelage, DL.; Low, S.; Rao, MA.; Hong, JS.; Little, TD. Family violence, bullying, fighting, and substance use among adolescents: A longitudinal mediational model. Journal of Research on Adolescence. 2013. http://dx.doi.org/10.1111/jora.12060

Farmer CA, Aman MG. Development of the children's scale of hostility and aggression: Reactive/proactive (C-SHARP). Research in Developmental Disabilities: A Multidisciplinary Journal. 2009; 30(6):1155–1167. http://dx.doi.org/10.1016/j.ridd.2009.03.001.

Farrington, D. Understanding and preventing bullying. In: Tonry, M., editor. Crime and justice. Chicago, IL: University of Chicago Press; 1993. p. 381-458.

Felix ED, Sharkey JD, Green JG, Furlong MJ, Tanigawa D. Getting precise and pragmatic about the assessment of bullying: the development of the California bullying victimization scale. Aggressive Behavior. 2011; 37(3):234–247. http://dx.doi.org/10.1002/ab.20389. [PubMed: 21404281]

Ferguson CJ, San Miguel C, Kilburn JC, Sanchez P. The effectiveness of school-based anti-bullying programs: A meta-analytic review. Criminal Justice Review. 2007; 32(4):401–414. http://dx.doi.org/10.1177/0734016807311712.

Finger, LF.; Yeung, AS.; Craven, RG.; Parada, RH.; Newey, K. Adolescent peer relations instrument: Assessment of its reliability and construct validity when use with upper primary students. Paper presented at the Australian Association for Research in Education; Australia: Brisbane; 2008.

Fitzpatrick S, Bussey K. The development of the social bullying involvement scales. Aggressive Behavior. 2011 Mar-Apr;37(2):177–192. http://dx.doi.org/10.1002/ab.20379. [PubMed: 21274854]

Furlong MJ, Sharkey JD, Bates MP, Smith DC. An examination of the reliability, data screening procedures, and extreme response patterns for the Youth Risk Behavior Surveillance Survey. Journal of School Violence. 2004; 3(2–3):109–130. http://dx.doi.org/10.1300/J202v03n02_07.

Furlong, MJ.; Sharkey, JD.; Felix, ED.; Tanigawa, D.; Green, J. Bullying assessment: A call for increased precision of self-reporting procedures. In: Jimerson, SR.; Swearer, SB.; Espelage, DL., editors. Handbook of bullying in schools: An international perspective. New York: NY: Routledge; 2010. p. 329-345.

Gable RK, Ludlow LH, McCoach BD, Kite SL. Development and validation of the survey of knowledge of internet risk and internet behavior. Educational and Psychological Measurement. 2011; 71(1):217–230. http://dx.doi.org/10.1177/0013164410387389.

Gini G, Pozzoli T. Association between bullying and psychosomatic problems: A meta-analysis. Pediatrics. 2009; 123(3):1059–1065. http://dx.doi.org/10.1542/peds.2008-1215. [PubMed: 19255040]

Gladden, RM.; Vivolo-Kantor, AM.; Hamburger, ME.; Lumpkin, C. Bullying surveillance: Uniform definitions and recommended data elements, Version 1.0. Atlanta, GA: 2014.

Goossens FA, Olthof T, Dekker PH. New participant role scales: Comparison between various criteria for assigning roles and indications for their validity. Aggressive Behavior. 2006; 32(4):343–357. http://dx.doi.org/10.1002/ab.20133.

Gottheil NF, Dubow EF. The interrelationships of behavioral indices of bully and victim behavior. Journal of Emotional Abuse. 2001a; 2(2–3):75–93. http://dx.doi.org/10.1300/J135v02n02_06.

Gottheil NF, Dubow EF. Tripartite beliefs models of bully and victim behavior. Journal of Emotional Abuse. 2001b; 2(2–3):25–47. http://dx.doi.org/10.1300/J135v02n02_03.

Greene, MB. Bullying and harassment in schools. In: Moser, RS.; Frantz, CE., editors. Shocking violence. Springfield, IL: Charles Thomas; 2000. p. 72-101.

Grief JL, Furlong MJ. The assessment of school bullying: Using theory to inform practice. Journal of School Violence. 2006; 5:33–50. http://dx.doi.org/10.1300/J202v05n03_04.

Hamburger, M.; Basile, K.; Vivolo, A. Measuring bullying victimization, perpetration, & bystander experiences: A compendium of assessment tools. Atlanta, GA: 2011.

Haynie DL, Nansel T, Eitel P, Crump AD, Saylor K, Yu K, et al. Bullies, victims, and bully/victims: Distinct groups of at-risk youth. The Journal of Early Adolescence. 2001; 21(1):29–49. http://dx.doi.org/10.1177/0272431601021001002.

Hinduja S, Patchin JW. Bullying, cyberbullying, and suicide. Archives of Suicide Research. 2010; 14(3):206–221. http://dx.doi.org/10.1080/13811118.2010.494133. [PubMed: 20658375]

Hunt C, Peters L, Rapee RM. Development of a measure of the experience of being bullied in youth. Psychological Assessment. 2012; 24:156–165. http://dx.doi.org/10.1037/a0025178. [PubMed: 22250594]

Ireland JL. Bullying behaviors among male and female prisoners: A study of adult and young offenders. Aggressive Behavior. 1999; 25:161–178. http://dx.doi.org/10.1002/(SICI)1098-2337(1999)25:3<161::AID-AB1>3.0.CO;2-#.

Kann L, Kinchen S, Shanklin SL, Flint KH, Hawkins J, et al. Youth risk behavior surveillance—United States, 2013. MMWR Surveillance Summaries. 2014; 63(SS-04):1–168.

Kert AS, Codding RS, Tryon GS, Shiyko M. Impact of the word “bully” on the reported rate of bullying behavior. Psychology in the Schools. 2010; 47(2):193–204.

Konishi C, Hymel S. Bullying and stress in early adolescence: The role of coping and social support. The Journal of Early Adolescence. 2009; 29:333–356. http://dx.doi.org/10.1177/02724316083201263.

Land D. Teasing apart secondary students’ conceptualizations of peer teasing, bullying and sexual harassment. School Psychology International. 2003; 24(2):147–165. http://dx.doi.org/10.1177/0143034303024002002.

Landis JR, Koch GG. The measurement of observer agreement for categorical data. Biometrics. 1977; 33(1):159–174. http://dx.doi.org/10.2307/2529310. [PubMed: 843571]

Leff, SS.; Power, TJ.; Goldstein, AB. Outcome measures to assess the effectiveness of bullying prevention programs in the schools. In: Espelage, DL.; Swearer, SB., editors. Bullying in American schools: A social–ecological perspective on prevention and intervention. 2004. p. 269-294.

Leff SS, Thomas DE, Shapiro ES, Paskewich B, Wilson K, Necowitz-Hoffman B, et al. Developing and validating a new classroom climate observation assessment tool. Journal of school violence. 2011; 10(2):165–184. http://dx.doi.org/10.1080/15388220.2010.539167. [PubMed: 21643447]

McConville D, Cornell D. Attitudes toward aggression and aggressive behavior among middle school students. Journal of Emotional and Behavioral Disorders. 2003; 11:179–187.

Mynard H, Joseph S. Development of the multidimensional peer-victimization scale. Aggressive Behavior. 2000; 26:169–178. http://dx.doi.org/10.1002/(SICI)1098-2337(2000)26:2<169::AID-AB3>3.3.CO;2-1.

Nakamoto J, Schwartz D. Is peer victimization associated with academic achievement? A meta-analytic review. Social Development. 2009; 19(2):221–242. http://dx.doi.org/10.1111/j.1467-9507.2009.00539.x.

Nansel TR, Craig W, Overpeck MD, Saluja G, Ruan W. Cross-national consistency in the relationship between bullying behaviors and psychosocial adjustment. Archives of Pediatrics & Adolescent Medicine. 2004; 158(8):730. http://dx.doi.org/10.1001/archpedi.158.8.730. [PubMed: 15289243]

Nishina A, Juvonen J. Daily reports of witnessing and experiencing peer harassment in middle school. Child development. 2005; 76(2):435–450. http://dx.doi.org/10.1111/j.1467-8624.2005.00855.x. [PubMed: 15784092]

Olweus, D. Bullying at School: What we know and what we can do. Malden, MA: Blackwell; 1993.

Orpinas P, Frankowski R. The aggression scale. The Journal of Early Adolescence. 2001; 21(1):50–67. http://dx.doi.org/10.1177/0272431601021001003.

Orpinas P, Horne AM, Staniszewski D. School bullying: Changing the problem by changing the school. School Psychology Review. 2003; 32:431–444.

Perry DG, Kusel SJ, Perry LC. Victims of peer aggression. Developmental Psychology. 1988; 24:807–814. http://dx.doi.org/10.1037//0012-1649.24.6.807.

Peters MP, Bain SK. Bullying and victimization rates among gifted and high-achieving students. Journal for the Education of the Gifted. 2011; 34(4):624–643. http://dx.doi.org/10.1177/016235321103400405.

Poteat VP, Espelage DL. Exploring the relation between bullying and homophobic verbal content: The homophobic content agent target (HCAT) scale. Violence and Victims. 2005; 20(5):513–528. http://dx.doi.org/10.1891/vivi.2005.20.5.513. [PubMed: 16248488]

Potter WJ, Levine-Donnerstein D. Rethinking validity and reliability in content analysis. Journal of Applied Communication Research. 1999; 27:258–284. http://dx.doi.org/10.1080/00909889909365539.

Prati G. Development and psychometric properties of the homophobic bullying scale. Educational and Psychological Measurement. 2012; 72:649–664. http://dx.doi.org/10.1177/0013164412440169.

Raine A, Dodge K, Loeber R, Gatzke-Kopp L, Lynam D, Reynolds C, et al. The reactive–proactive aggression questionnaire: Differential correlates of reactive and proactive aggression in adolescent boys. Aggressive Behavior. 2006; 32:159–171. http://dx.doi.org/10.1002/ab.20115. [PubMed: 20798781]

Rigby, K. What it takes to stop bullying in schools: An examination of the rationale and effectiveness of school-based interventions. In: Furlong, MJ.; Bates, MP.; Smith, DC.; Kingery, P., editors. Appraisal and prediction of school violence. Hauppauge, NY: Nova Science; 2004. p. 165-191.

Rigby K, Slee PT. Dimensions of interpersonal relation among Australian children and implications for psychological well-being. The Journal of Social Psychology. 1993; 133(1):33–42. http://dx.doi.org/10.1080/00224545.1993.9712116. [PubMed: 8464217]

Rivers I, Poteat VP, Noret N, Ashurst N. Observing bullying at school: The mental health implications of witness status. School Psychology Quarterly. 2009; 24(4):211–223. http://dx.doi.org/10.1037/a0018164.

Salazar, LF.; DiClemente, RJ.; Crosby, RA. Measurement and design related to theoretically-based public health research and practice. In: Salazar, LF.; DiClemente, RJ.; Crosby, RA., editors. Health behavior theory for public health: Principles, foundations, and applications. Burlington, MA: Jones & Bartlett Publishers; 2011. p. 255-285.

Salmivalli C, Voeten M. Connections between attitudes, group norms, and behaviors associated with bullying in schools. International Journal of Behavioral Development. 2004; 28:246–258.

Schäfer M, Korn S, Smith PK, Hunter SC, Mora-Merchán JA, Singer MM, et al. Lonely in the crowd: Recollections of bullying. British Journal of Developmental Psychology. 2004; 22:379–394. http://dx.doi.org/10.1348/0261510041552756.

Smith PK, Cowie H, Olafsson RF, Liefooghe A. Definitions of bullying: A comparison of terms used, and age and gender differences, in a fourteen-country international comparison. Child development. 2002; 73(4):1119–1133. http://dx.doi.org/10.1111/1467-8624.00461. [PubMed: 12146737]

Smith, PK.; del barrio, C.; Tokunaga, RS. Definitions of bullying and cyberbullying: How useful are the terms?. In: Bauman, S.; Cross, D.; Walker, J., editors. Principles of cyberbullying research: Definitions, measures and methodology. New York: NY: Routledge; 2013. p. 26-40.

Smith, PK.; Thompson, D. Practical approaches to bullying. London, UK: David Fulton Publishers; 1991.

Solberg ME, Olweus D. Prevalence estimation of school bullying with the Olweus bully/victim questionnaire. Aggressive Behavior. 2003; 29(3):239–268. http://dx.doi.org/10.1002/ab.10047.

Spector, PE. Summated rating scale construction: An introduction. Sage; 1992.

Swearer SM, Cary PT. Perceptions and attitudes toward bullying in middle school youth. Journal of Applied School Psychology. 2003; 19(2):63–79. http://dx.doi.org/10.1300/J008v19n02_05.

Swearer, SM.; Siebecker, AB.; Johnsen-Frerichs, LA.; Wang, C. Assessment of bullying/victimization: The problem of comparability across studies and across methodologies. In: Jimerson, SR.; Swearer, SB.; Espelage, DL., editors. Handbook of bullying in schools: An international perspective. New York: NY: Routledge; 2010. p. 305-327.

Tarshis TP, Huffman LC. Psychometric properties of the peer interactions in primary school (PIPS) questionnaire. Journal of Developmental & Behavioral Pediatrics. 2007; 28(2):125–132. http://dx.doi.org/10.1097/01.DBP.0000267562.11329.8f. [PubMed: 17435463]

Taub J. Evaluation of the second step violence prevention program at a rural elementary school. School Psychology Review. 2002; 31(2):186–200.

Thompson JK, Fabian LJ, Moulton DO, Dunn ME, Altabe MN. Development and validation of the physical appearance related teasing scale. Journal of Personality Assessment. 1991; 56(3):513–521. http://dx.doi.org/10.1207/s15327752jpa5603_12. [PubMed: 1865309]

Thompson JK, Cattarin J, Fowler B, Fisher E. The perception of teasing scale (POTS): A revision and extension of the physical appearance related teasing scale (PARTS). Journal of Personality Assessment. 1995; 65:146–157. http://dx.doi.org/10.1207/s15327752jpa6501_11. [PubMed: 16367650]

Tokunaga RS. Following you home from school: A critical review and synthesis of research on cyberbullying and victimization. Computers in Human Behavior. 2010; 26:277–287. http://dx.doi.org/10.1016/j.chb.2009.11.014.

Vaillancourt T, McDougall P, Hymel S, Krygsman A, Miller J, Stiver K, et al. Bullying: Are researchers and children/youth talking about the same thing? International Journal of Behavioral Development. 2008; 32(6):486–495. http://dx.doi.org/10.1177/0165025408095553.

Van Schoiack-Edstrom L, Frey KS, Beland K. Changing adolescents' attitudes about relational and physical aggression: An early evaluation of a school-based intervention. School Psychology Review. 2002; 31(2):201–216.

Veenstra R, Lindenberg S, Oldehinkel AJ, De Winter AF, Verhulst FC, Ormel J. Bullying and victimization in elementary schools: a comparison of bullies, victims, bully/victims, and uninvolved preadolescents. Developmental psychology. 2005; 41(4):672. http://dx.doi.org/10.1037/0012-1649.41.4.672. [PubMed: 16060813]

Vernberg EM, Jacobs AK, Hershberger SL. Peer victimization and attitudes about violence during early adolescence. Journal of Clinical Child Psychology. 1999; 28(3):386–395. [PubMed: 10446688]

Vivolo A, Holt M, Massetti G. Individual and contextual factors for bullying and peer victimization: Implications for prevention. Journal of School Violence. 2011; 10:201–212. http://dx.doi.org/10.1080/15388220.2010.539169.

Warden D, Cheyne B, Christie D, Fitzpatrick H, Reid K. Assessing children's perceptions of prosocial and antisocial peer behaviour. Educational Psychology. 2003; 23(5):547–567. http://dx.doi.org/10.1080/0144341032000123796.

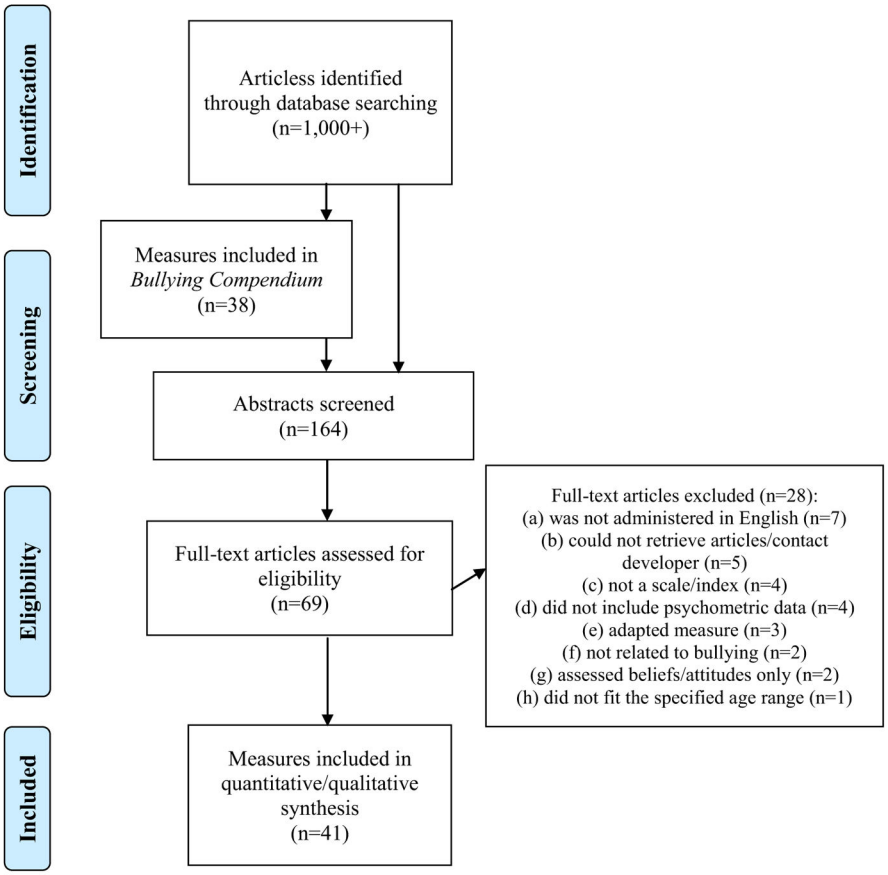

Fig. 1.
PRISMA flow of information through the different phases of a systematic review.

Table 1

Characteristics of the 41 included measures.

Variables	Mean (Range)	Count (Percent)
Sample size	1089.4 (47–8693)	
Age range[a]	10.5–15.7 (3–25)	
Grade level range[b]	5.9–8.9 (0–16)	
Number of items	27.4 (5–96)	
Publication year	2003.4 (1988–2012)	
Study location[c]		
U.S.		19 (46.3)
International		15 (36.5)
Multiple countries		3 (7.3)
Used term "bullying" in measure		
Yes		13 (31.7)
No		28 (68.3)
Provided participants a definition of bullying		
Yes		11 (26.8)
No		21 (51.2)
Unknown		9 (22)
Components within definition[d,e]		
Power imbalance		9 (81.8)
Intention to cause harm		9 (81.8)
Victim experiences harm		6 (66.7)
Repetition		7 (63.6)
Aggressive behaviors		8 (72.7)
Reporting method[e]		
Youth self-report only		31 (75.6)
Peer nomination only		5 (12.2)
Combination of youth self-report and peer nomination		2 (4.9)
Teacher report		2 (4.9)
Parent report		1 (2.4)
Response options[e]		
Binary (e.g., yes/no, true/false)		11 (26.8)
Scale/index		33 (80.5)
Open-ended (includes peer nomination)		10 (24.4)
Multiple choice		2 (4.9)
Time frame		
Past 7 days/week		4 (9.8)
Past 30 days/month		9 (22)
Past 2 to 3 months or school semester		3 (7.3)
Past 4 months		1 (2.4)

Variables	Mean (Range)	Count (Percent)
Current school year		2 (4.9)
Past 12 months/year		1 (2.4)
Other (e.g., recently or when growing up)		4 (9.8)
Unknown		17 (41.4)
Type of bullying behavior measured[e]		
Perpetration-only		3 (7.3)
Victimization-only		8 (19.5)
Both perpetration and victimization		29 (70.8)
Bystander with either perpetration or victimization		1 (2.4)
All three: perpetration, victimization, and bystander		7 (17.1)

[a] 22 measures provided information on age range or average age.

[b] 32 measures provided information on grade levels. A zero for grade level denotes kindergarten and “16” is senior in college.

[c] 37 articles provided information on study location.

[d] n = 11

[e] Categories are not mutually exclusive.

Table 2

Behavior content of the 41 included measures.

Behavior content	Count (Percent)
Making fun, teasing, embarrassing	33 (80.5)
Physical acts	32 (78.0)
Calling names	30 (73.2)
Making threats	25 (61.0)
Rude comments and gestures	23 (56.1)
Social exclusion	23 (56.1)
Stealing or damaging property	18 (43.9)
Spreading rumors	15 (36.6)
Cyber-bullying	10 (24.4)
Sexual harassment	8 (19.5)
Legal harassment	7 (17.1)
Group bullying	7 (17.1)
Bystander or witness	7 (17.1)
Other broad behaviors	6 (14.6)
Weight-based teasing	6 (14.6)
Homophobic teasing	2 (4.9)
Weapon-carrying	2 (4.9)

Table 3

Definition characteristics and measured components of included measures.

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
Adolescent Peer Relations Instrument (Finger, Yeung, Craven, Parada, & Newey, 2008)	36	Victimization: α = 0.89; perpetration: α = 0.82	No	n/a			X	X	
Aggression Scale (Orpinas & Frankowski, 2001)	11	Perpetration: α = 0.88	No	n/a		X	X	X	
Bull-S Questionnaire (Cerezo & Ato, 2005)	25	Victimization: α = 0.83; Perpetration: α = 0.82	No	n/a				X	
Bullying-Behavior Scale (Austin & Joseph, 1996)	12	Victimization: α = 0.82; perpetration: α = 0.83	No	n/a			X	X	
California Bullying Victimization Scale (Felix et al., 2011)	12	κ for 8 items = 0.46–0.64; κ for bully/victim = 0.71	No	n/a	X	X	X	X	X
Child Social Behavior Questionnaire (Warden, Cheyne, Christie, Fitzpatrick, & Reid, 2003)	53	Peer nomination: α = 0.90; teacher report: α = 0.90; self-report: α = 0.68	No	n/a		X	X	X	X
Children's Scale of Hostility and Aggression: Reactive/ Proactive (C-SHARP) (Farmer & Aman, 2009)	58	Verbal aggression: α = 0.92; bullying: α = 0.89; covert aggression: α = 0.88; physical aggression: α = 0.74	No	n/a			X	X	

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
Children's Social Behavior Scale — Self Report (Crick & Grotpeter, 1995)	15	Overt aggression: α = 0.94; relational aggression: α = 0.83	No	n/a		X	X	X	
Cyber Victim & Bullying Scale (Cetin, Yaman, & Peker, 2011)	44	Victimization: α = 0.89, split half = 0.79, test–retest = 0.85; perpetration: α = 0.89, split half = 0.79, test–retest = 0.90	No	n/a			X	X	
Direct and Indirect Patient Behavior Checklist — Prisoner Version (Ireland, 1999)	87	Direct bullying: α = 0.91; verbal bullying: α = 0.72; physical bullying: α = 0.78; indirect bullying: α = 0.77; theft: α = 0.85; sexual harassment: α = 0.79	No	n/a	X		X	X	X
Gatehouse Bullying Scale (Bond, Wolfe, Tollit, Butler, & Patton, 2007)	12	Any bullying: rho test–retest = 0.65, κ= 0.63	No	n/a		X	X		X
Homophobic Bullying Scale (Prati, 2012)	42	Homophobic aggression towards gay men: α = 0.82; homophobic aggression towards lesbian women: α = 0.87; homophobic victimization: α = 0.86; witnessing homophobic aggression	No	n/a		X	X	X	

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
		towards gay men: α = 0.82; witnessing homophobic aggression towards lesbian women: α = 0.83							
Homophobic Content Agent Target Scale (Poteat & Espelage, 2005)	10	Victimization: α = 0.85; perpetration: α = 0.85	Yes	None			X	X	
Illinois Bully Scale (Espelage & Holt, 2001)	18	Victimization: α = 0.88; perpetration: α = 0.87; physical fighting: α = 0.83	No	n/a	X		X	X	
Introducing My Classmates (Gottheil & Dubow, 2001a)	8	n/a	No	n/a		X		X	
Modified Aggression Scale (Bosworth, Espelage, & Simon, 1999)	5	Perpetration: α = 0.83	No	n/a		X	X	X	
Modified Peer Nomination Inventory (Perry, Kusel, & Perry, 1988)	14	Victimization: α = 0.96; split-half alpha range = 0.78–0.98; test–retest r = 0.93 over 3 months	No	n/a		X	X		
Multidimensional Peer-Victimization Scale (Mynard & Joseph, 2000)	16	Physical victimization: α = 0.85; verbal victimization: α = 0.75; social manipulation: α = 0.77; attacks on	Yes	Power imbalance Intent to cause harm		X	X	X	

Aggress Violent Behav. Author manuscript; available in PMC 2016 January 06.

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
		property: α = 0.73 = 0.73							
New Participant Role Scale (Goossens, Olthof, & Dekker, 2006)	32	Perpetrator: α = 0.96; follower: α = 0.95; outsider: α = 0.94; defender: α = 0.88; victim: α = 0.93	No	n/a				X	X
Olweus Bully/ Victim Questionnaire (Solberg & Olweus, 2003)	36	Victimization: α = 0.88; perpetration: α = 0.87	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors Victim experiences harm	X	X	X	X	X
Pacific Rim Bullying Measure (Konishi & Hymel, 2009)	12	Victimization: α = 0.73; perpetration: α = 0.77	Yes	Power imbalance Repetition Aggressive behaviors		X	X	X	
Participant Role Questionnaire (Salmivalli & Voeten, 2004)	15	Perpetration: α = 0.93; assistant α = 0.95; reinforcer α = 0.90; defender α = 0.89; outsider α = 0.88	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors Victim experiences harm		X	X	X	
Peer Interactions in Primary School Questionnaire (Tarshis & Huffman, 2007)	22	Total α = 0.90; victimization: test–retest r = 0.87; perpetration: test–retest r = 0.76	No	n/a		X	X	X	X
Peer Relations Assessment Questionnaire (Rigby & Slee, 1993)	20	Victimization α = 0.86 (School A), α = 0.78 (School B); perpetration α = 0.75 (School A), α = 0.78 (School B)	No	n/a	X	X	X	X	

Aggress Violent Behav. Author manuscript; available in PMC 2016 January 06.

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
Peer Relationship Survey (Cho, Hendrickson, & Mock, 2009)	20	Victimization: α = 0.89; perpetration: α = 0.90	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors Victim experiences harm		X	X		
Perception of Teasing Scale (Thompson et al., 1995)	22	Weight-teasing victimization α = 0.88, test–retest for frequency = 0.90, test–retest for effect = 0.85; competence-teasing victimization α = 0.75, test–retest for frequency = 0.82, test–retest for effect = 0.66	No	n/a			X	X	X
Personal Experiences Checklist (Hunt, Peters, & Rapee, 2012)	32	Overall test–retest r = 0.79 (0.61– 0.86); verbal–relational bullying: α = 0.91, test–retest r = 0.75; cyber bullying α = 0.90, test–retest r = 0.86; physical bullying α = 0.91, test–retest r = 0.61; bullying based on culture α = 0.78, test–retest r = 0.77	No	n/a			X	X	
Physical Appearance Related Teasing Scale (Thompson, Fabian, Moulton,	18	Weight-size teasing victimization α = 0.91, test–retest = 0.86;	No	n/a			X		

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
Dunn, & Altabe, 1991)		general appearance teasing victimization α = 0.71, test–retest = 0.87							
Reactive–Proactive Aggression Questionnaire (Raine et al., 2006)	23	Reactive perpetration: α = 0.84; proactive perpetration: α = 0.86; total perpetration: α = 0.90	No	n/a	X	X	X	X	X
Retrospective Bullying Questionnaire (Schäfer et al., 2004)	44	Primary school victimization: test– retest r = 0.88; secondary school victimization test–retest r = 0.87	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors			X	X	
Reynolds Bully-Victimization Scale for Schools (Peters & Bain, 2011)	46	Victimization: α = 0.93, test–retest r = 0.80; perpetration: α = 0.93, test–retest r = 0.81	No	n/a	X	X	X	X	X
School Climate and Bullying Scale (McConville & Cornell, 2003)	59	Victimization: α = 0.75	Yes	Power imbalance Intent to cause harm Aggressive behaviors Victim experiences harm	X	X	X	X	X
Self Report Inventory of Setting the Record Straight (Gottheil & Dubow, 2001b)	15	Victimization: α = 0.88; perpetration: α = 0.72	No	n/a		X		X	
Social Bullying Involvement Scales (Fitzpatrick & Bussey, 2011)	96	Social victim: α = 0.97; social bully α = 0.93; social witness: α = 0.96; social intervener: α = 0.97	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors Victim experiences harm		X	X	X	X

Measure	Total items	Reliability	Definition used	Definition components	Components measured[a]				
					Power imbalance	Intent to cause harm	Repetition	Aggressive behaviors	Victim experiences harm
Survey of Knowledge of Internet Risk & Internet Behavior (Gable, Ludlow, McCoach, & Kite, 2011)	10	Total: α = 0.69	No	n/a		X		X	
The School Life Survey (Chan, Myron, & Crawshaw, 2005)	24	Victimization α = 0.83, test–retest r = 0.939; perpetration test–retest r = 0.835	No	n/a		X	X	X	
The Swearer Bully Survey (Swearer & Cary, 2003)	31	Total: α = 0.87; physical bullying: α = 0.79; verbal bullying: α = 0.85	Yes	Power imbalance Intent to cause harm Repetition Aggressive behaviors	X	X	X	X	X
Traditional Bullying and Cyberbullying Survey (Hinduja & Patchin, 2010)	34	Victimization: α = 0.88; perpetration: α = 0.88; cyber victimization: α = 0.74; cyber perpetration: α = 0.76	No	n/a		X	X	X	X
Victimization of Self and Victimization of Others (Vernberg, Jacobs, & Hershberger, 1999)	18	Victimization: α = 0.85; perpetration: α = 0.78	Yes	Intent to cause harm Victim experiences harm		X	X	X	X
Victimization Scale (Orpinas, Horne, & Staniszewski, 2003)	10	Victimization: α = 0.86	No	n/a		X	X	X	X
Weight-Based Teasing Scale (Eisenberg, Neumark-Sztainer, & Perry, 2003)	5	Victimization: α = 0.84	No	n/a			X	X	

[a]This column represents the item-specific language or response option language. This should be distinguished from the "definition components" column, which describes which components were included in the bullying definition provided to participants.